Which model to use ?

Microlens modellers face many dilemmas:

Blending or no blending?
Keep original error bars, or resacle them?
Simple scale factor, or more elaborate model.
Point source or extended source?
Free or fix limb darkening?
Include parallax? Include xallarap?
Single or binary lens?
Single or binary source?
Companion at d or 1/d ?
Low q or high q companion?

How can we deal with these decisions automatically?

A similar but simpler example:

Fit N = *30* points with
M = *1, 2, ... 20* polynomial coefficients.

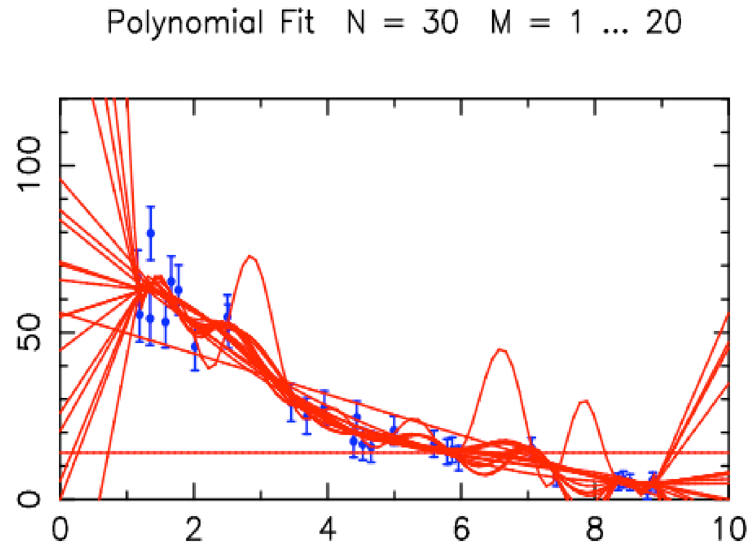
"Badness-of-fit" statistic:

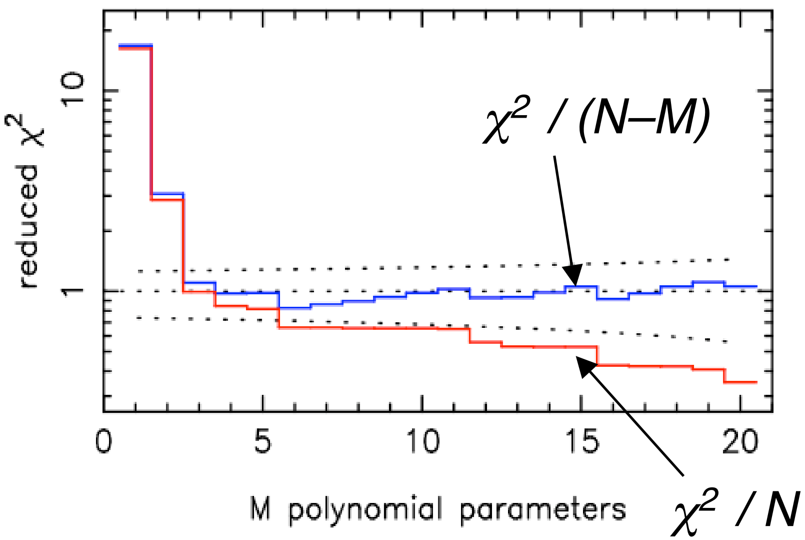
$$\frac{\chi^2}{N-M} \approx 1 \pm \sqrt{\frac{2}{N-M}}$$

χ^2 rejects M = *1, 2,*
accepts M = *3, 4, ...*

Higher M = more flexible model, lower χ^2, but less satisfactory fit.

χ^2 is not the whole story.





Bayes: parameter estimation

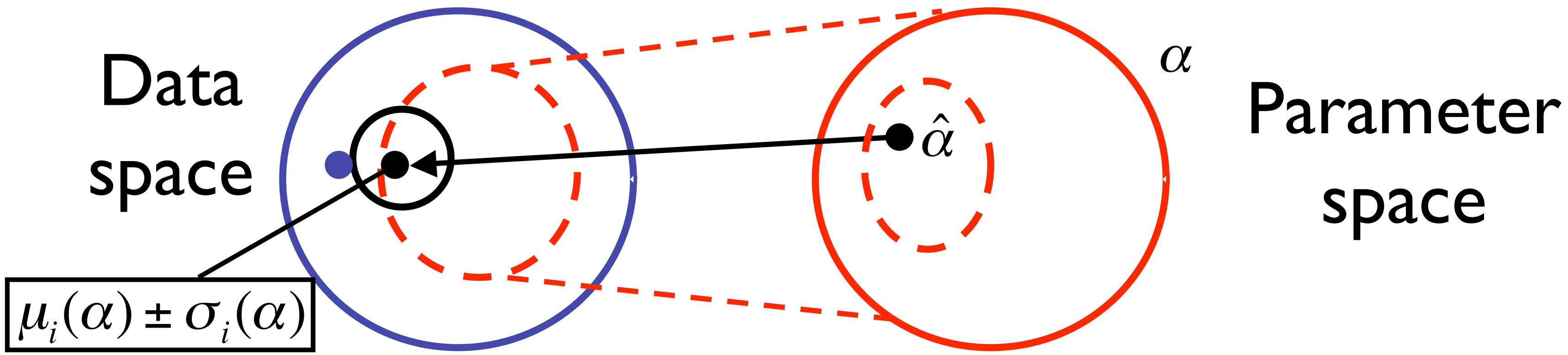
D = data M = model α = parameters

$$P(D \mid \alpha, M) = \frac{L_D(\alpha, M)}{Z_D(\alpha, M)} \qquad \text{Likelihood}: \; L_D(\alpha, M) = \exp\left\{-\frac{\chi^2(D,\alpha,M)}{2}\right\}$$

$$\text{Data Volume}: \; Z_D(\alpha, M) \equiv \int L_D(\alpha, M)\, dD = (2\pi)^{N/2} \prod_{i=1}^{N} \sigma_i$$

$$\text{Posterior}: \; P(\alpha \mid D, M) = \frac{P(D \mid \alpha, M)\, P_M(\alpha)}{\int P(D \mid \alpha, M)\, P_M(\alpha)\, d\alpha} = \frac{L_D(\alpha, M)}{Z_D(\alpha, M)} \frac{P_M(\alpha)}{Z_M(D)}$$

$$\text{Prior}: \; P_M(\alpha) \qquad \text{Support}: \; Z_M(D) \equiv \int \frac{L_D(\alpha, M)\, P_M(\alpha)}{Z_D(\alpha, M)}\, d\alpha = \left\langle \frac{L_D}{Z_D} \right\rangle_{P_M}$$



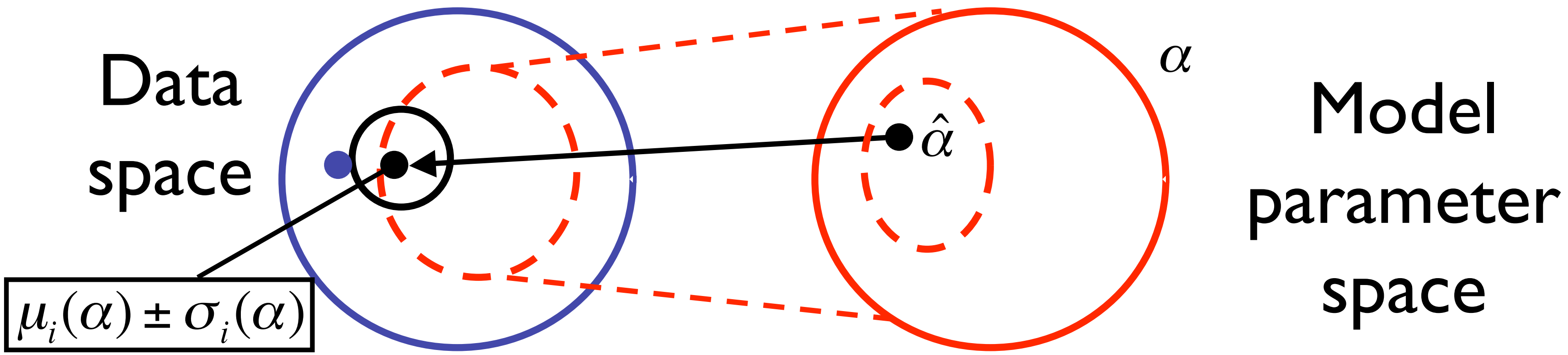
Bayes: parameter estimation

$$P(\alpha \mid D) = \frac{P(D \mid \alpha)\, P(\alpha)}{\int P(D \mid \alpha)\, P(\alpha)\, d\alpha} = \frac{L_D(\alpha)}{Z_D(\alpha)} \frac{P(\alpha)}{Z(D)} = \frac{L_D(\alpha)/Z_D(\alpha)}{\langle L_D / Z_D \rangle} P(\alpha)$$

$$\text{Likelihood}: \; L_D(\alpha) = e^{-\chi^2/2} \quad \text{Data Volume}: \; Z_D(\alpha) = (2\pi)^{N/2} \prod \sigma_i$$

$$\text{Prior}: \; P(\alpha) \qquad \text{Support}: \; Z(D) \equiv \int \frac{L_D(\alpha)\, P(\alpha)}{Z_D(\alpha)}\, d\alpha = \langle L_D / Z_D \rangle$$

1. $L_D(\alpha)$ large => Model fits the data well.
2. $Z_D(\alpha)$ small => Model has small error bars.
3. $P(\alpha)$ large => Model is not "too crazy".
4. $Z(D)$ support for the model (explain later)



Bayes: parameter estimation

Parameter uncertainties:

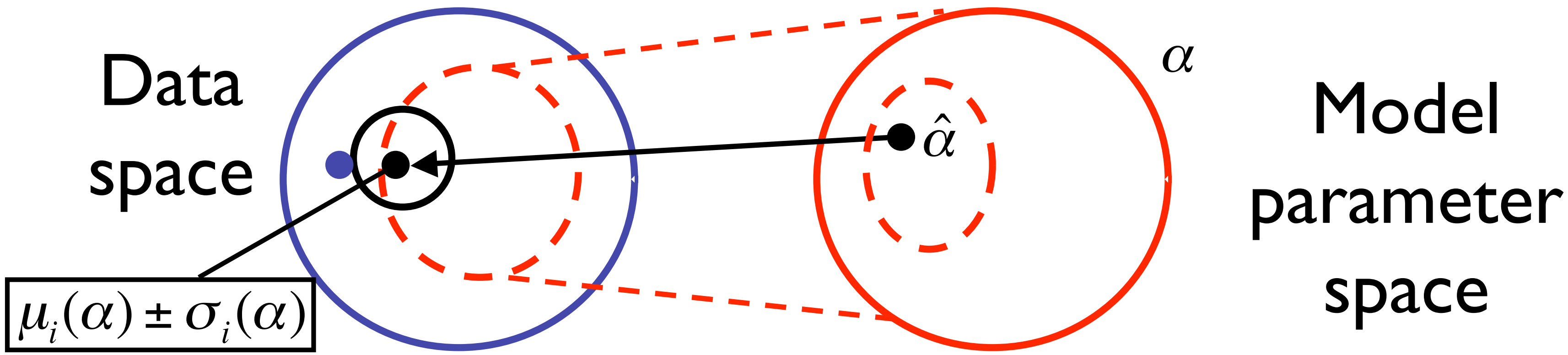
1. $L_D(\alpha)$ Fit the data.
2. $Z_D(\alpha)$ Small error bars.
3. $P_M(\alpha)$ Be not "too crazy".
4. $Z_M(D)$ same for all α.

Posterior distribution:

$$P(\alpha \mid D) = \frac{L_D(\alpha)}{Z_D(\alpha)} \frac{P(\alpha)}{Z(D)}$$

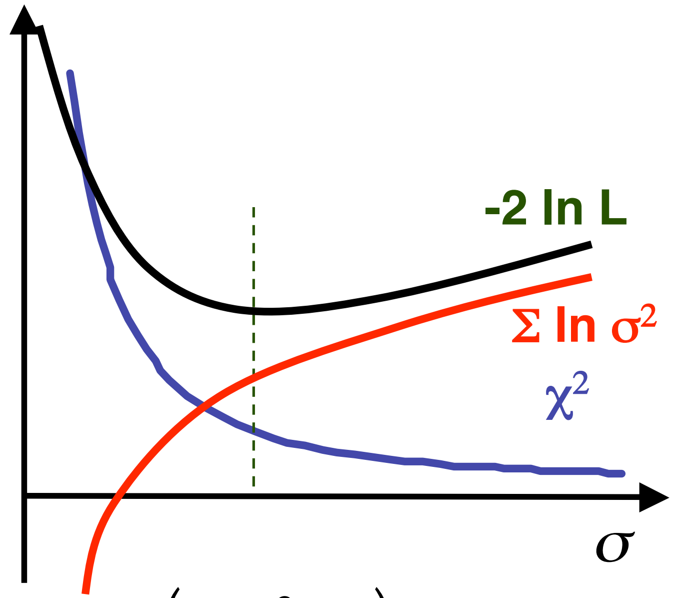
Ratio of posterior probability densities:

$$\frac{P(\alpha_1 \mid D)}{P(\alpha_2 \mid D)} = \left(\frac{L_D(\alpha_1)}{L_D(\alpha_2)} \Big/ \frac{Z_D(\alpha_1)}{Z_D(\alpha_2)} \right) \frac{P(\alpha_1)}{P(\alpha_2)} \quad = \exp\left\{ -\frac{\Delta\chi^2}{2} \right\} \left(\prod_{i=1}^{N} \frac{\sigma_i(\alpha_2)}{\sigma_i(\alpha_1)} \right) \frac{P(\alpha_1)}{P(\alpha_2)}$$



Penalty for large error bars:

$$\chi^2 = \sum_{i=1}^{N} \frac{(D_i - \mu_i)^2}{\sigma_0^{\ 2} + f^2 \sigma_i^{\ 2}}$$



χ^2 minimisation fails :

$$\chi^2 \to 0 \quad \text{as } \sigma \to \infty$$

Maximum likelihood :

$$L(\text{model}) = P(\text{data} \mid \text{model}) = \frac{\exp\left(-\chi^2/2\right)}{(2\pi)^{N/2} \prod_{i=1}^{N} \left(\sigma_0^{\ 2} + f^2 \sigma_i^{\ 2}\right)^{1/2}}$$

$$-2 \ln L = \chi^2 + \sum_i \ln\left(\sigma_0^{\ 2} + f^2 \sigma_i^{\ 2}\right) + N \ln(2\pi)$$

Comparing Models

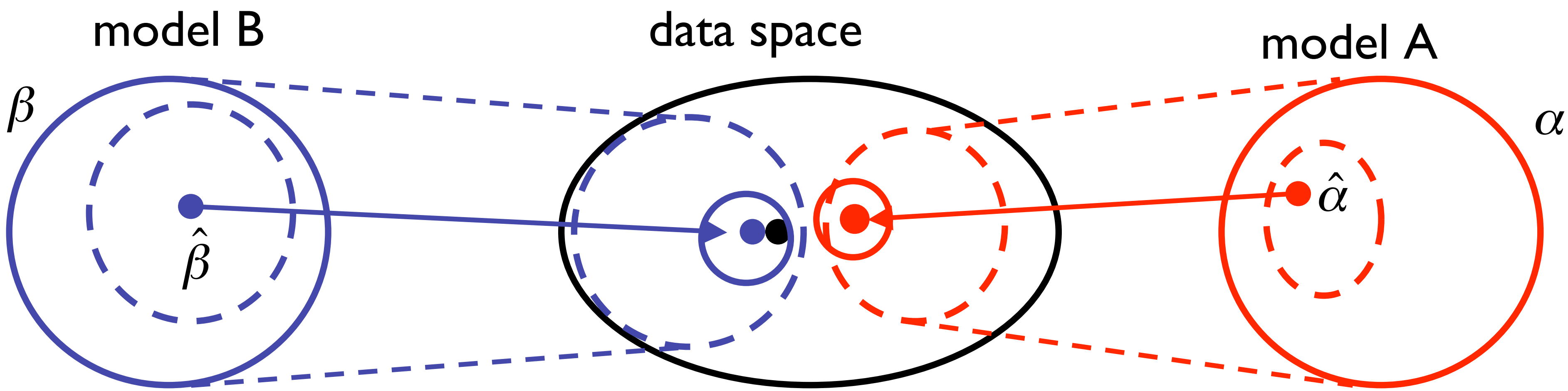
$$P(D \mid \alpha, M) = \frac{L_D(\alpha, M)}{Z_D(\alpha, M)} = \frac{e^{-\chi^2/2}}{(2\pi)^{N/2}\prod \sigma_i} \qquad Z_M(D) \equiv \int \frac{L_D(\alpha, M)\, P_M(\alpha)}{Z_D(\alpha, M)}\, d\alpha$$

Posterior : joint distribution over parameters α and models M :

$$P(\alpha, M \mid D) = \frac{P(D \mid M,\alpha)\, P(\alpha \mid M)\, P(M)}{\int P(D \mid M,\alpha)\, P(\alpha \mid M)\, P(M)\, d\alpha\, dM} = \frac{L_D(\alpha, M)}{Z_D(\alpha, M)} \frac{P_M(\alpha)\, P(M)}{Z(D)}$$

Posterior distribution on models (integrate over parameters α) :

$$P(M \mid D) = \int P(\alpha, M \mid D)\, d\alpha = \frac{P(M)}{Z(D)} \int \frac{L_D(\alpha, M)}{Z_D(\alpha, M)} P_M(\alpha)\, \mathrm{d}\alpha = \frac{Z_M(D)}{Z(D)} P(M)$$

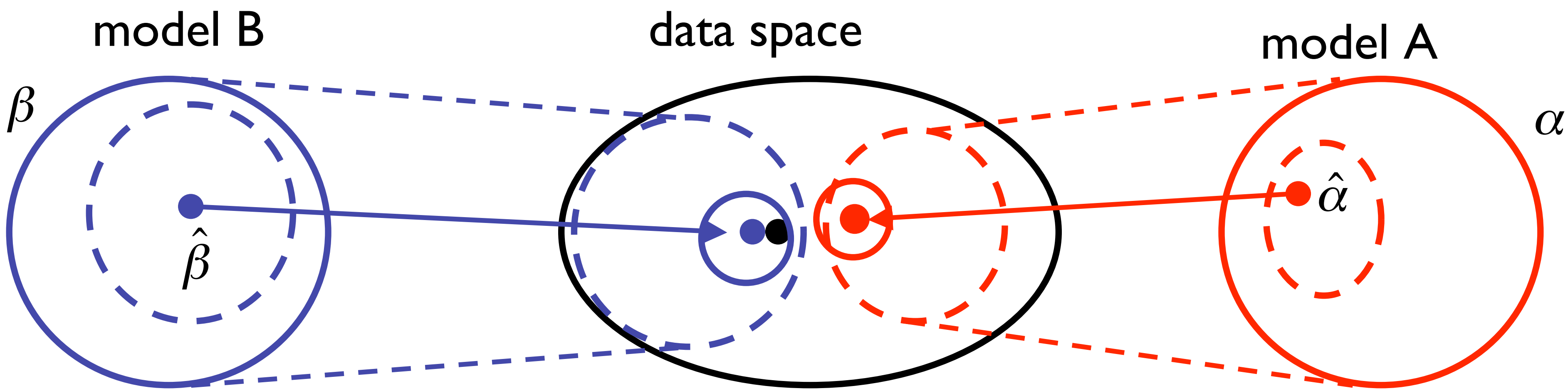


Comparing Models

Posterior distribution: $P(M \mid D) = \dfrac{Z_M(D)}{Z(D)} P(M)$

$$\frac{P(A \mid D)}{P(B \mid D)} = \frac{Z_A(D)}{Z_B(D)} \frac{P(A)}{P(B)} = \frac{\langle L_D / Z_D \rangle_A}{\langle L_D / Z_D \rangle_B} \frac{P(A)}{P(B)}$$

Support for model M: $Z_M(D) \equiv \left\langle \dfrac{L_D}{Z_D} \right\rangle_{P_M} = \displaystyle\int \frac{L_D(\alpha)\, P_M(\alpha)}{Z_D(\alpha)}\, d\alpha$



Comparing Models

Support for model M: $\quad Z_M(D) \equiv \int \frac{L_D(\alpha)\, P_M(\alpha)}{Z_D(\alpha)}\, d\alpha$

$$Z_M(D) \approx \frac{L_D(\hat{\alpha})\, P_M(\hat{\alpha})}{Z_D(\hat{\alpha})} \prod_{k=1}^{M} \left(2\pi\,\sigma^2(\hat{\alpha}_k)\right)^{1/2} = \frac{e^{-(\chi^2_{\min}/2)} \prod_{k=1}^{M} \sigma(\hat{\alpha}_k)}{\left(2\pi\right)^{(N-M)/2} \prod_{i=1}^{N} \sigma_i}\, P_M(\hat{\alpha})$$

1. Penalty for bad fit: $\quad L_D(\hat{\alpha}) = e^{-(\chi^2_{\min}/2)}$

2. Penalty for large error bars: $\quad \frac{1}{Z_D(\hat{\alpha})} = \frac{1}{\left(2\pi\right)^{N/2} \prod_{i=1}^{N} \sigma_i}$

3. Penalty for fine tuning: $\quad P_M(\hat{\alpha})\left(2\pi\right)^{M/2} \prod_{k=1}^{M} \sigma(\hat{\alpha}_k)$

(posterior parameter volume
/ prior parameter volume).

A possible approach:

1. Use Galaxy model, adjusting parameters to fit observed OGLE/MOA event parameter distributions, to establish priors.
2. Run competing models in parallel.
3. Use MCMC with to track evolving parameters of each model.
4. Use MCMC to evaluate the relative probability of different models.